

PRINCIPAL COMPONENT ANALYSIS UNTUK ANALISA POLA TANGKAPAN IKAN DI INDONESIA

Titin Kristiana

Program Studi Manajemen Informatika AMIK BSI Jakarta

Jl. Kramat raya no. 18 Jakarta Pusat

email: titin.tka@bsi.ac.id

Abstrak

Tangkapan ikan yang beragam jenis di Indonesia sangat banyak diketahui ada lebih dari 80 jenis ikan yang tertangkap di perairan Indonesia. Untuk mengetahui informasi jenis ikan yang tertangkap diperlukan analisa terhadap pola data hasil tangkapan ikan sehingga dapat diketahui jenis ikan apa saja yang tertangkap. Pencarian pola atau hubungan asosiatif dari data yang berskala besar sangat erat kaitannya dengan data mining. Analisis asosiasi atau *association rule mining* adalah teknik data mining untuk menemukan aturan asosiatif antara suatu kombinasi item. Dalam metode *association rule* terdapat dua proses yaitu proses pembangkitan *Frequent Itemset* dan penggalian kaidah asosiasi. Pembangkitan *Frequent Itemset* merupakan sebuah proses untuk mendapatkan itemset yang saling berhubungan dan memiliki nilai asosiasi berdasarkan nilai support dan confidence. Adapun Algoritma yang digunakan untuk membangkitkan frequent itemset ini adalah Algoritma Apriori. Algoritma apriori memiliki kelemahan pada ekstraksi fitur yang sesuai pada atribut yang digunakan sehingga menyebabkan rule yang terbentuk menjadi banyak. dalam penelitian ini menerapkan algoritma apriori berbasis *principal component analysis* untuk mendapatkan rule yang lebih optimal. Setelah dilakukan *eksperiment* menggunakan algoritma apriori dengan besaran $\Phi = 30$, min Support 80% dan min Confidence 80% hasil rule yang terbentuk berjumlah 82 rules. Sedangkan *eksperiment* kedua yang dilakukan dengan menggunakan algoritma apriori berbasis *principal component analysis* dengan besaran $\Phi = 30$, min Support 80% dan min Confidence 80% hasil yang terbentuk berjumlah 12 rules dengan seluruhnya *lift ratio* 1

Kata kunci: Penangkapan Ikan, Algoritma Apriori, *Association rule*, *principal component analysis*

Abstract

Different kinds of fish in Indonesia is very much known to exist more than 80 species of fish caught in the waters of Indonesia. To find out which type of fish caught necessary analysis of the data pattern catches so as to know what kind of fish are caught. Search pattern or associative relationships of large-scale data that are closely related to data mining. Analysis of the association or the association rule mining is a data mining technique to discover the rules of associative between a combination of items. In the association rule method, there are two processes, namely the process of generating *Frequent Itemset* and trenching association rules. *Frequent Itemset Generation* is a process to get itemset interconnected and has a value of association based on the value of support and confidence. The algorithm used to generate the frequent itemset is Apriori Algorithm. Apriori algorithm has a weakness in the appropriate feature extraction that is used to attribute causing rule that formed a research banyak. dalam berbasis applying apriori algorithm principal component analysis to obtain a more optimal rule. After experiments using apriori algorithm with a magnitude $\Phi = 30$, min Support 80% and 80% Confidence min rule formed results totaled 82 rules. While the second experiment was done by using an algorithm based on principal component analysis prior the magnitude $\Phi = 30$, min Support 80% and 80% Confidence min formed results amounted to 12 rules to fully lift the ratio of 1

Keywords: Fish catches, Apriori algorithm, Association rule, Principle component analysis

I. PENDAHULUAN

Indonesia sebagai Negara kepulauan memiliki potensi perikanan yang sangat besar dan beraneka ragam, Indonesia memiliki 17.508 pulau dengan garis pantai sepanjang 104.000 km dan 70% (5,8 juta km²) dari luas Indonesia adalah lautan. (PUSDATIN, 2011). Pemanfaatan sumber daya kelautan yang tersimpan di wilayah perairan nusantara sampai saat ini perlu mendapat perhatian yang memadai mengingat besarnya potensi yang tersimpan dalam sumber daya laut. Bila dibanding dengan luas wilayah lautan yang dimiliki Indonesia maka usaha pemerintah dalam memanfaatkan potensi kelautan masih membutuhkan banyak peningkatan. Sumber daya perikanan laut merupakan salah satu aspek penting dari sumber daya laut nusantara yang perlu digali dan dikembangkan. Potensi perikanan yang ada di perairan Indonesia diketahui melimpah, mengingat perairan Indonesia merupakan daerah tropis dengan kandungan plankton yang kaya, sebagai tempat berkumpulnya ikan-ikan.

Penangkapan ikan adalah kegiatan yang bertujuan untuk memperoleh ikan di perairan yang tidak dalam keadaan dibudidayakan dengan alat atau cara apapun, termasuk kegiatan yang menggunakan kapal untuk membuat, mengangkut, menyimpan, mendinginkan, atau mengawetkan (Mulyana, 2008). Hasil tangkapan ikan yang beragam jenis di Indonesia sangat banyak diketahui ada lebih dari 80 jenis ikan yang tertangkap di perairan Indonesia. Untuk mengetahui informasi jenis ikan yang tertangkap diperlukan analisa terhadap pola data hasil tangkapan ikan sehingga dapat diketahui jenis ikan apa saja yang tertangkap.

Pencarian pola atau hubungan asosiatif dari data yang berskala besar sangat erat kaitannya dengan data mining. Analisis asosiasi atau *association rule mining* adalah teknik data mining untuk menemukan aturan asosiatif antara suatu kombinasi item. Salah satu tahap analisis asosiasi adalah untuk menghasilkan algoritma yang efisien dengan analisis pola frekuensi tertinggi (*frequent pattern mining*) (Kusrini, 2009), dalam metode *association rule* terdapat dua proses yaitu proses pembangkitan *Frequent Itemset* dan penggalan kaidah asosiasi. Pembangkitan *Frequent Itemset* merupakan sebuah proses untuk mendapatkan itemset yang saling berhubungan dan memiliki nilai asosiasi berdasarkan nilai support dan confidence. Adapun Algoritma yang digunakan untuk

membangkitkan frequent itemset ini adalah Algoritma Apriori.

Association rule dengan menggunakan Algoritma Apriori untuk analisa pola data hasil tangkapan ikan (Tyas, 2008), *Mining gene expression databases for association rules* (Hanash, 2003), *Feature Selection for Large Scale Rule Mining and Information Gain: a Hybrid Approach* (Balamurugan P. r., 2009)

Pada penelitian ini mengusulkan sebuah metode filter untuk menemukan atribut asosiasi yang lebih efektif menggunakan *principal component analysis* digunakan untuk menyederhanakan variable yang diamati dengan cara menyusutkan (mereduksi) dimensinya.

Algoritma apriori dapat menyelesaikan masalah penentuan itemset dan aturan dalam data besar yang ada pada data tangkapan ikan. tetapi algoritma apriori memiliki kelemahan pada ekstraksi fitur yang sesuai pada data dan atribut yang digunakan sehingga menyebabkan rule yang terbentuk menjadi banyak

Tujuan penelitian ini adalah menerapkan metode *principal component analysis* untuk ekstraksi fitur pada data tangkapan ikan sehingga rule yang terbentuk dapat berkurang dalam algoritma apriori untuk menentukan pola tangkapan ikan di Indonesia

II. KAJIAN LITERATUR

Perikanan adalah semua kegiatan yang berhubungan dengan pengelolaan dan pemanfaatan sumber daya ikan dan lingkungannya mulai dari praproduksi, produksi, pengolahan sampai dengan pemasaran yang dilaksanakan dalam suatu system bisnis perikan sedangkan sumber daya ikan adalah potensi semua jenis ikan. Penangkapan ikan adalah kegiatan yang bertujuan untuk memperoleh ikan di perairan yang tidak dalam keadaan dibudidayakan dengan alat atau cara apapun, termasuk kegiatan yang menggunakan kapal untuk membuat, mengangkut, menyimpan, mendinginkan, atau mengawetkan (Siombo, 2010).

Berdasarkan tempat hidupnya jenis ikan dibagi menjadi 2 kategori, sebagai berikut:

1. Ikan Pelagis merupakan ikan yang hidup dilapisan air bagian atas
2. Ikan Demersal merupakan ikan yang biasa hidup didasar perairan

Data mining adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dengan menggunakan teknik pengenalan pola seperti teknik statistic dan matematika (Larose, 2007)

Data Mining merupakan teknologi baru yang sangat berguna untuk membantu perusahaan-perusahaan menemukan informasi yang sangat penting dari gudang data mereka. Beberapa aplikasi data mining fokus pada prediksi, mereka meramalkan apa yang akan terjadi dalam situasi baru dari data yang menggambarkan apa yang terjadi di masa lalu (Witten, Frank, & Hall, 2011)

Kemajuan luar biasa yang terus berlanjut dalam bidang *data mining* didorong oleh beberapa faktor, antara lain (Larose, 2007):

1. Pertumbuhan yang cepat dalam kumpulan data.
2. Penyimpanan data dalam *data warehouse*, sehingga seluruh perusahaan memiliki akses ke dalam *database* yang handal.
3. Adanya peningkatan akses data melalui navigasi *web* dan *intranet*. Tekanan kompetisi bisnis untuk meningkatkan penguasaan pasar dalam globalisasi ekonomi.
4. Perkembangan teknologi perangkat lunak untuk *data mining* (ketersediaan teknologi).
5. Perkembangan yang hebat dalam kemampuan komputasi dan pengembangan kapasitas media penyimpanan.

Association rule (aturan asosiasi) atau seringkali disebut juga dengan *association analysis* merupakan suatu studi tentang “apa bersama apa”, yang pada awalnya merupakan studi tentang *database* transaksi pelanggan untuk menentukan kebiasaan suatu produk dibeli bersama dengan produk apa, maka aturan asosiasi juga sering di sebut dengan *market basket analysis*.

Aturan asosiasi dapat memberikan informasi dalam bentuk hubungan “*if-then*” atau “jika-maka”. Aturan ini di hitung dari data yang sifatnya probabilistik (Santosa, 2007) Analisis asosiasi di definisikan, merupakan suatu proses untuk menemukan semua aturan asosiatif yang memenuhi syarat minimum untuk *support* (*minimum support*) dan syarat minimum untuk *confidence* (*minimum confidence*), yang nantinya berguna dalam menentukan *interesting association rules*, yaitu untuk dibandingkan dengan batasan (*threshold*) yang ditentukan oleh *user*. Bila

memenuhi kedua batasan, maka sebuah *rule* dapat disebut *interesting rule*.

Disebut *support* karena ia mengukur seberapa tingkat dukungan data terhadap validitas aturan yang dikembangkan, *support* biasanya dinyatakan dalam ukuran prosentase (%).

$$\text{Support}(A) = \frac{\text{Jumlah Transaksi mengandung } A}{\text{Total Transaksi}}$$

Sementara itu, nilai *support* dari 2 item diperoleh dari rumus sebagai berikut:

$$\text{Support}(A, B) = P(A \cap B)$$

$$= \frac{\text{Jumlah transaksi mengandung } A \text{ dan } B}{\text{Total Transaksi}}$$

- *Confidence* adalah rasio antara jumlah transaksi yang meliputi semua *item* dalam *antecedent* dan *consequent* dengan jumlah transaksi yang meliputi semua *item* dalam *antecedent*, atau dapat di tulis:

$$\text{Confidence} = P(B|A)$$

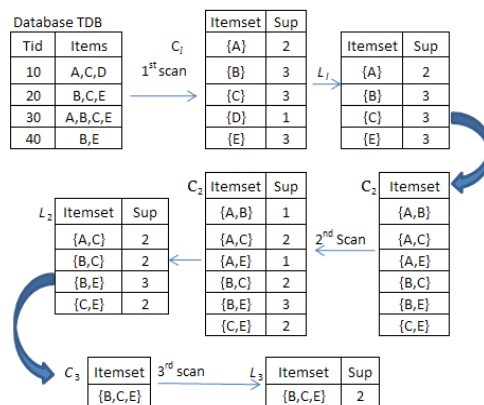
$$= \frac{\text{Jumlah transaksi mengandung } A \text{ dan } B}{\text{Total Transaksi mengandung } A}$$

Algoritma apriori pertama kali diusulkan oleh Agrawal, Imielinski dan Swami dalam “*Mining Association Rules between Sets of Item in Large Databases*”, yang mengolah data transaksi dalam suatu *database* dengan mencari kombinasi *item*. Kemudian mencari seluruh kaidah apriori dalam aturan asosiasi berdasarkan nilai *support* dan *confidence*. Algoritma apriori bertujuan untuk menemukan *frequent itemset* yang dijalankan oleh sekumpulan data. Pada iterasi ke-k, akan ditemukan semua *itemset* yang memiliki k *item*, yang disebut dengan k-*itemset*.

Faktor-faktor yang dapat mempengaruhi kompleksitas pada algoritma apriori adalah sebagai berikut:

1. Pemilihan *minimum support*
 - Dengan menurunkan batas *minimum support* dapat menyebabkan semakin banyaknya *frequent itemset* yang didapatkan.
 - Hal tersebut juga akan menyebabkan peningkatan jumlah dari kandidat dan panjang maksimum *frequent itemset*.
2. Dimensi atau jumlah *item* pada *data set*.
 - Lebih banyak ruang yang dibanyak ruang yang dibutuhkan untuk menyimpan *hitungan support* untuk setiap *item*.

- Jika jumlah pada *frequent item* juga meningkat, baik komputasi dan I/O *cost* juga mungkin akan meningkat.
- 3. Besarnya ukuran *database*.
 - Karena apriori membuat *multiple pass*, *run time* dari algoritma juga akan meningkat dengan jumlah dari transaksi.
- 4. Rata-rata panjang transaksi
 - Lebar transaksi akan meningkatkan kepadatan *data set*.
 - Hak ini akan meningkatkan panjang maksimum dari *frequent itemset* dan garis lintang pada *hash tree* (jumlah dari subset di dalam sebuah transaksi meningkatkan lebarnya).



Gambar 1. Contoh dataset

Feature Extraction merupakan langkah preprocessing yang menghilangkan varians dari dataset, sehingga pengklasifikasi hilir atau estimator regresi tampil lebih baik (Maimon, 2010). Pengurangan Dimensi dari data ke ruang dimensi yang lebih rendah yang terkait erat (*reductionis* sering dimensi digunakan sebagai langkah dalam ekstraksi fitur), tetapi tujuan bisa berbeda. Reduksi dimensi memiliki sejarah panjang sebagai metode untuk visualisasi data, dan untuk mengekstraksi fitur kunci dimensi rendah menunjukkan bahwa, di bawah asumsi keteraturan tertentu, tingkat optimal convergence¹ untuk regressionvaries nonparametrik seperti $m-p / (2p + d)$ Di mana m adalah ukuran sampel, data terletak pada R^d , dan di mana fungsi regresi diasumsikan p kali terdiferensialkan.. Pertimbangkan 10.000 sampel poin, untuk $p = 2$ dan $d = 10$. Jika d meningkat menjadi 20, jumlah titik sampel harus meningkat menjadi sekitar 10 juta untuk mencapai tingkat optimal konvergensi yang sama. Jika data kebohongan kami (sekitar) pada L berjenis rendah dimensi yang terjadi untuk dimasukkan dalam H berjenis tinggi

dimensi, pemodelan data proyeksi L daripada di H dapat mengubah masalah tidak layak menjadi layak satu.

Dalam ekstraksi fitur dan pengurangan dimensi terbagi menjadi 2 kategori besar, yaitu:

1. Metode yang bergantung pada proyeksi

Metode ini merupakan pendekatan yang paling sederhana untuk menemukan proyeksi dimensi rendah yang mengekstrak informasi yang berguna dari data, dengan memaksimalkan fungsi tujuan yang cocok. Metode ini adalah dasar dari *projection pursuit*, dimana proyeksi optimal ditemukan dalam proyeksi sebelumnya. Selain menangani data dimensi tinggi, metode proyeksi mengejar dapat mengukur kuatnya fitur bising atau tidak relevan.

Jika data terdistribusi secara normal, menemukan proyeksi bersama yang data berkorelasi setara dengan menemukan proyeksi bersama yang bersifat independen, sehingga meskipun menggunakan analisis komponen utama akan cukup untuk menemukan proyeksi independen, menemukan proyeksi sepanjang mana data statistik independen adalah lebih kuat daripada menemukan proyeksi sepanjang data tersebut berkorelasi. Oleh karena itu ICA berkonsentrasi pada situasi di mana distribusi data berangkat dari normalitas, dan pada kenyataannya, menemukan maksimal non-Gaussian komponen akan memberikan sebuah komponen independen.

2. Metode yang mencoba untuk model manifold di mana data terletak metode untuk pengurangan dimensi yang mengasumsikan bahwa data yang terletak pada manifold dimensi rendah tertanam dalam ruang H tinggi dimensi, yaitu embedding lokal linier, skala multidimensi, Isomap, Laplacian eigenmaps, dan spektral clustering.

Principal Component Analysis pada dasarnya bertujuan untuk menyederhanakan variabel yang diamati dengan cara menyusutkan (mereduksi) dimensinya. Hal ini dilakukan dengan cara menghilangkan korelasi diantara variabel bebas melalui transformasi variabel bebas asal ke variabel baru yang tidak berkorelasi sama sekali (Maimon, 2010).

III. METODE PENELITIAN

Pada metode penelitian eksperimen, terdapat beberapa tahapan penelitian yang dilakukan seperti berikut:

a. Pengumpulan data

Pada bagian ini dijelaskan tentang bagaimana dan darimana data dalam penelitian ini didapatkan, ada dua tipe dalam pengumpulan data, yaitu pengumpulan data primer dan pengumpulan data sekunder. Data primer adalah data yang dikumpulkan pertama kali untuk melihat apa yang sesungguhnya terjadi. Data sekunder adalah data yang sebelumnya pernah dibuat oleh seseorang baik di terbitkan atau tidak (Kothari, 2004). Pengumpulan data primer bisa didapat dari model interview terhadap responden, maupun dengan model observasi terhadap suatu badan yang sedang diteliti (Gray, 2004). Pada tahap ini ditentukan data yang akan diproses. Mencari data yang tersedia, memperoleh data tambahan yang dibutuhkan, mengintegrasikan semua data kedalam data set, termasuk variabel yang diperlukan dalam proses.

b. Pengolahan awal data

Pada bagian ini dijelaskan tentang tahap awal data mining. Pengolahan awal data meliputi proses input data ke format yang dibutuhkan, pengelompokan dan penentuan atribut data, serta pemecahan data (split) untuk digunakan dalam proses pembelajaran (training) dan pengujian (testing).

c. Model yang diusulkan

Pada tahap ini data dianalisis, dikelompokkan variabel mana yang berhubungan dengan satu sama lainnya. Setelah data dianalisis lalu diterapkan model-model yang sesuai dengan jenis data. Pembagian data kedalam data latihan (*training data*) dan data uji (*testing data*) juga diperlukan untuk pembuatan model.

d. Eksperimen dan pengujian model

Pada bagian ini dijelaskan tentang langkah-langkah eksperimen meliputi cara pemilihan arsitektur yang tepat dari model atau metode yang diusulkan sehingga didapatkan hasil yang dapat membuktikan bahwa metode yang digunakan adalah tepat.

e. Evaluasi dan validasi hasil

Pada bagian ini dilakukan evaluasi dan validasi hasil penerapan terhadap model penelitian yang dilakukan untuk mengetahui tingkat keakurasian model.

Jumlah data awal yang diperoleh dari pengumpulan data yaitu sebanyak 1.200 data, Untuk mendapatkan data yang berkualitas, data dibersihkan dan ditransformasikan kebentuk yang diinginkan sehingga dapat dilakukan persiapan dalam pembuatan model. Untuk itu diperlukan teknik untuk mem-filter data training atau dikenal dengan teknik *preprocessing* (Witten, Frank, & Hall, 2011). Dimana hasil dari teknik *preprocessing* menghasilkan atribut yang berbeda. Berikut tabel hasil teknik *preprocessing* (Han & Kamber, 2007) yaitu:

- a. Data *cleansing* digunakan untuk membersihkan nilai yang kosong, tidak konsisten atau ada data yang kosong (*missing value* dan *noisy*).
- b. Data *integration* digunakan untuk menyatukan tempat penyimpanan yang berbeda-beda ke dalam satu tempat penyimpanan yang sama. Dalam hal ini data diambil dari data *warehouse* yaitu data pelanggan dan data transaksi penjualan.
- c. Data *reduction*, dimana jumlah data tupel yang digunakan untuk data training terlalu besar dan hanya beberapa atribut yang diperlukan, sehingga atribut yang tidak perlu akan dihapus. Apabila terdapat tupel yang sama maka hanya menggunakan satu tupel saja.

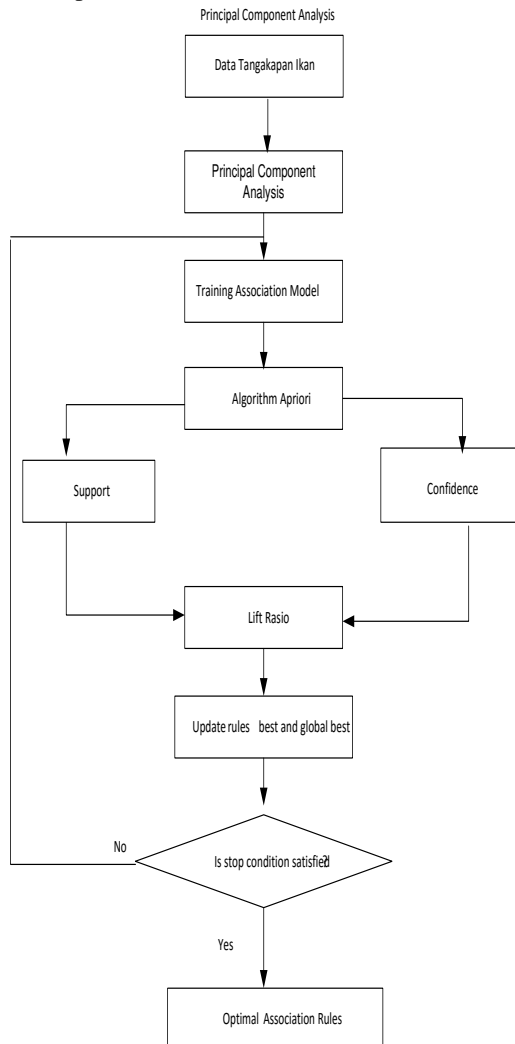
Tahap modeling untuk menyelesaikan analisa pola tangkapan ikan dengan menggunakan dua metode yaitu algoritma apriori dan algoritma *principal component analysis*

1. Algoritma Apriori yaitu metode untuk menemukan *frequent itemset* yang dijalankan oleh sekumpulan data
2. Principal Component Analysis (PCA) yaitu metode ekstraksi fitur untuk menyederhanakan variabel yang diamati dengan cara menyusutkan (mereduksi) dimensinya. Hal ini dilakukan dengan cara menghilangkan korelasi diantara variabel bebas melalui transformasi variabel bebas asal ke variabel baru yang tidak berkorelasi sama sekali

IV. PEMBAHASAN

Model yang diusulkan pada penelitian tentang analisa pola tangkapan ikan adalah dengan menerapkan algoritma apriori menggunakan *principal component analysis*. Penerapan algoritma apriori menentukan itemset dengan *support* dan *confidence*

terlebih dahulu. Kemudian didapatkan nilai rasio dari itemset yang terbentuk. Sedangkan algoritma apriori berbasis *principal component analysis* menentukan cumulative variance terbesar dari hasil transformasi sehingga terbentuk aturan yang lebih efisien untuk pemecahan masalah tersebut



Gambar 2. Model yang Diusulkan

Setelah melalui proses pembersihan ada 437 *record* data tangkapan ikan, dengan 31 jenis ikan yang akan dipakai dalam aturan asosiasi. Pada tabel dibawah inidapat dilihat 31 jenis ikan yang akan dipakai dalam aturan asosiasi.

Nama atribut yang digunakan dalam penelitian ini yaitu:

1. Beronang Kuning
2. Beronang lingkis
3. Biji angka

4. Gabus Laut
5. Gulamah/Tigawaja
6. ikan gaji
7. Ikan kakap merah/Bambangan
8. ikan lidah
9. Ikan sebelah (Terompa)
10. Ikan sebelah mata kiri
11. Jenaha
12. Kakap Batu
13. Kakap Merah
14. Kakap putih
15. Kapas-kapas
16. Kerapu bebek
17. Kerapu karang
18. Kerapu karet
19. Kuniran
20. Kurau
21. Kurisi
22. Kuro/senangin
23. Layur
24. Lencam
25. Manyung
26. Pari burung
27. Pari hidung
28. sekop
29. Pari kembang/Pari macan
30. Pari kekeh
31. Remang
32. Swanggi

Dari *data training* inilah yang akan dipakai untuk menganalisis data. Langkah-langkah algoritma pada *association rule* adalah:

1. Tentukan besarnya besaran Φ , besaran nilai minimum *support* dan nilai minimum *confidence*.
2. Tentukan semua *frekuensi itemset* sering.
3. Untuk setiap *frekuensi itemset* lakukan hal sbb:
 - a. Ambil sebuah unsur, namakanlah s
 - b. Untuk sisanya namakanlah ss-s
 - c. Masukkan unsur-unsur yang diumpamakan ke dalam *rule if (ss-s) then s*.

Dari *data training* yang akan dipakai, terlebih dahulu ditentukan:

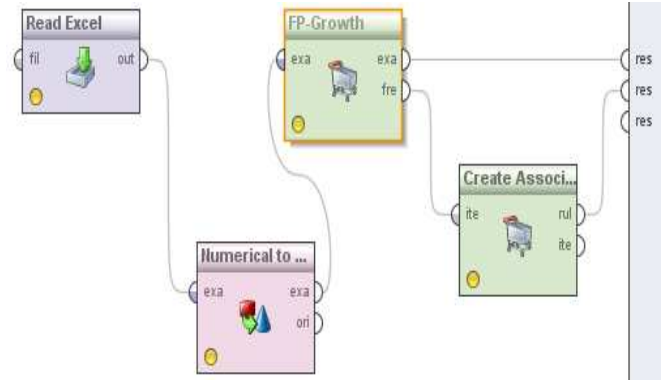
1. Mentukan besarnya:
 - a. besaran Φ , yaitu = 30
 - b. besarnya nilai minimum *support* = 80%
 - c. besarnya nilai minimum *confidence* = 80%

{Gulamah/Tigawaja, kakap putih, Layur},
 {Gulamah/Tigawaja, kakap putih, Pari Burung},
 {Gulamah/Tigawaja, Kapas-kapas, Kuniran},
 {Gulamah/Tigawaja, Kapas-kapas, Kurisi},
 {Gulamah/Tigawaja, Kerapu karang, Kuniran},
 {Gulamah/Tigawaja, Kerapu karang, Kurisi},
 {Gulamah/Tigawaja, Kerapu karang, Layur},
 {Gulamah/Tigawaja, Kerapu karang, Pari burung},
 {Gulamah/Tigawaja, Kuniran, Kurisi},
 {Gulamah/Tigawaja, Kuniran, Layur},
 {Gulamah/Tigawaja, Kuniran, Pari Burung},
 {Gulamah/Tigawaja, Kurisi, layur},
 {Gulamah/Tigawaja, Kurisi, Pari Burung},
 {Ikan Kakap merah/bambangan, Kurisi, layur},
 {Ikan Sebelah mata kiri, kakap putih, kapas-kapas}}

3. Rule yang dipakai adalah *if x then y*, dimana *x* adalah *antecedent* dan *y* adalah *consequent*. Berdasarkan rule tersebut, maka dibutuhkan dua buah item yang mana salah satunya sebagai *antecedent* dan sisanya sebagai *consequent*. Dari langkah kedua didapat 1 buah *Fk* yaitu *F2*. *F1* tidak disertakan karena hanya terdiri dari 1 item saja. Untuk *antecedent* boleh lebih dari 1 unsur, sedangkan untuk *consequent* terdiri dari 1 unsur.

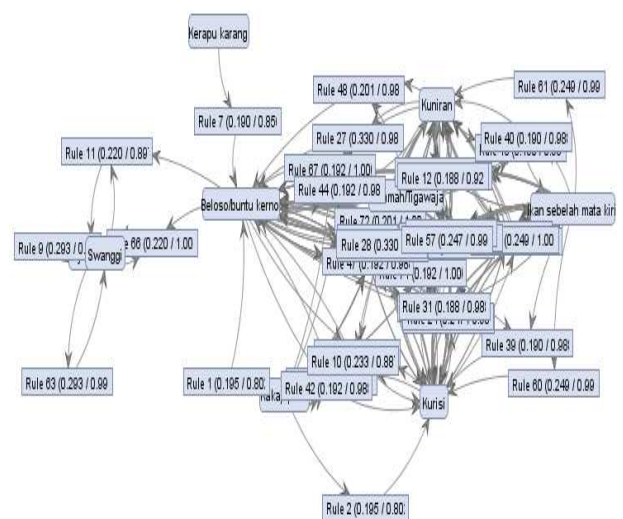
Hasil dari pengujian model yang dilakukan adalah analisa pola tangkapan ikan dengan algoritma apriori dan algoritma apriori berbasis *principal component analysis*. hasil eksperimen menggunakan algoritma apriori dengan besaran $\Phi = 30$, min Support 80% dan min Confidence 80% hasil rule yang terbentuk berjumlah 82 rules yang dapat dilihat pada gambar dibawah ini

Formula-formula pada algoritma Apriori sudah dirangkum dalam model algoritma Apriori pada framework RapidMiner. Formulanya sebagai berikut:



Gambar 3 Desain Model

Maka didapat hasil rule yang terbentuk menjadi 82 rule dari hasil rapidminer sebagai berikut :



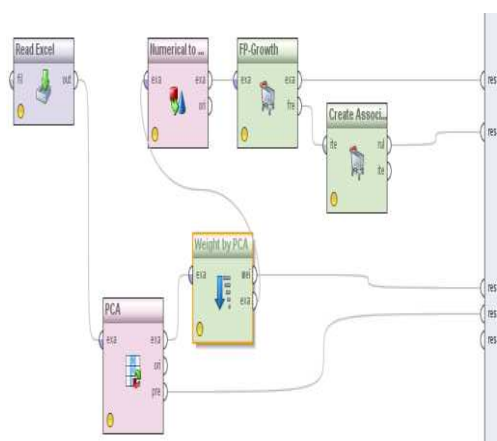
Gambar 4. Hasil Rule yang terbentuk

Setelah itu hasil eksperimen dengan penerapan ekstraksi fitur diterapkan menggunakan *principal component analysis* untuk efisiensi rule. Hasil

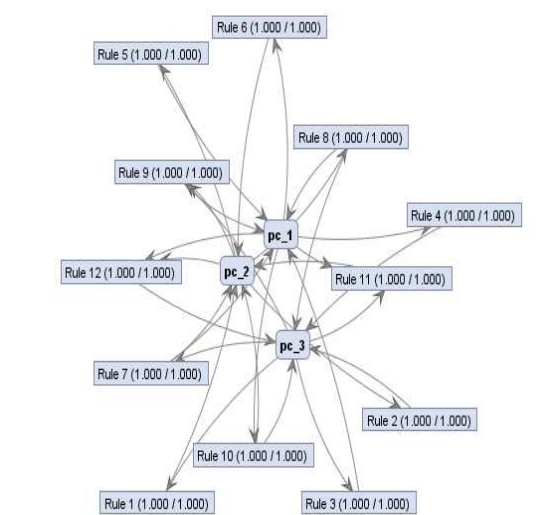
Maximizes Mutual Information on Gaussian Data, dengan entropy:

$$H(y) = - \int p(y) \log_2 p(y) dy = \frac{1}{2} \log_2 (e(2\pi)^d) + \frac{1}{2} \log_2 \det(C_y)$$

Sedangkan eksperimen kedua yang dilakukan dengan menggunakan algoritma apriori berbasis *principal component analysis* dengan besaran $\Phi = 30$, min Support 80% dan min Confidence 80% hasil yang terbentuk berjumlah 12 rules



Gambar 5. Model dengan PCA



Gambar 5. Hasil Rule setelah ekstraksi fitur

V. KESIMPULAN

Dalam penelitian ini menunjukkan hasil eksperimen yang dilakukan untuk analisa pola tangkapan ikan, dapat disimpulkan bahwa:

1. Hasil eksperimen menggunakan algoritma apriori dengan besaran $\Phi = 30$, min Support 80% dan min Confidence 80% hasil rule yang terbentuk berjumlah 82 rules
2. Sedangkan eksperimen kedua yang dilakukan dengan menggunakan algoritma apriori berbasis *principal component analysis* dengan besaran $\Phi = 30$, min Support 80% dan min Confidence 80% hasil yang terbentuk berjumlah 12 rules dengan seluruhnya *lift ratio* 1.

Maka dapat disimpulkan pengujian pengujian data tangkapan ikan menggunakan *algoritma apriori* berbasis *principal component analysis* lebih baik dari pada *algoritma apriori* sendiri. Dengan demikian dari hasil pengujian model diatas dapat disimpulkan bahwa algoritma apriori berbasis *principal component analysis* dapat memecahkan masalah dalam analisa pola tangkapan ikan

Dari hasil pengujian yang telah dilakukan dan hasil kesimpulan yang diberikan maka ada saran atau usul yang di berikan antara lain:

1. Data yang di pakai di harapkan menggunakan data yang terbaru, karena hasilnya dapat di pakai untuk strategi penangkapan ikan.
2. Agar lebih memperhatikan dimensi atau jumlah *item* pada *data set*, karena akan membutuhkan ruang yang banyak untuk penyimpanan hitungan *support* untuk setiap *item*.
3. Penelitian semacam ini dapat dikembangkan algoritma aturan asosiasi yang lain, seperti algoritma *quantitative association rule* dan algoritma *generalized association rule*

DAFTAR PUSTAKA

- Large Scale Data by Combining Class Association Rule Mining and Information : a Hybrid Approach. *Internetworking Indonesian Journal* , 17.
- Balamurugan, P. r. (2009). Feature Selection for Large Scale Rule Mining and Information Gain : a Hybrid Approach . *Internetworking Indonesia Journal* , 17-24.
- Berndtsson, M., Olsson, J. H., & Lundell, B. (2008). *Thesis Projects A Guide for Students in Computer Science and Information Systems*. Verlag London: Springer.
- Dawson, C. W. (2009). *Projects in Computing and Information System A Student's Guide*. England: Addison-Wesley.
- Gray, D. E. (2004). *Doing Research in the Real World*. New Delhi: SAGE.
- Han, J., & Kamber, M. (2007). *Data Mining Concepts and technique*. San Fransisco: Morgan Kaufmann.
- Hanash, C. C. (2003). Mining gene expresion databases for association rule. *Bioinformatics* , 79-86.

- K, G. S., & Deepa, D. S. (2011). Analysis of Computing Algorithm using Momentum in Neural Networks. *Journal of computing, volume 3, issue 6*, 163-166.
- Kothari, C. R. (2004). *Research Methology methodes and Technique*. India: New Age Interntional.
- Kusrini, E. t. (2009). *Algoritma Data Mining*. Yogyakarta: Andi.
- Larose, D. T. (2007). *Data Mining Methods And Models*. New Jersey: A John Wiley & Sons.
- Maimon, O. (2010). *Data Mining and Knowledge Discovery Handbook*. New York Dordrecht Heldelberg London: Springer.
- PUSDATIN. (2011). *Kelautan dan Perikanan dalam angka 2010*. Jakarta: Kementerian kelautan dan Perikanan.
- Santosa, B. (2007). *Data Mining Teknik Pemanfaatan Data Untuk keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- Siombo, M. r. (2010). *Hukum perikanan nasional dan internasional*. Jakarta: Gramedia Pustaka Utama.
- Tyas, E. W. (2008). Penerapan Association rule dengan menggunakan algoritma apriori untuk analisa pola data hasil tangkapan ikan. *Konferensi dan temu Teknologi Informasi dan Komunikasi Untuk Infonesia*.
- Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining Practical Machine Learning Tools and Techniques Third Edition*. USA: Morgan Kaufmann.
- Yaya Mulyana, A. D. (2008). *Konservasi kawasan perairan indonesia bagi masa depan dunia*. Jakarta: Direktorat Konservasi dan Taman Nasional laut, Kementerian Kelautan dan Perikanan.

Titin Kristiana, lahir di Jakarta, 31 Januari 1983. Pendidikan terakhir Megister Ilmu Komputer di STMIK Nusamandiri Jakarta lulus tahun 2012. Sejak Maret tahun 2008 sampai sekarang bekerja di Lembaga Pendidikan Bina Sarana Informatika sebagai dosen ilmu komputer. Selain mengajar saya juga membimbing mahasiswa dalam penyusunan tugas akhir.